Children's identification of unfamiliar voices on both target-present and target-absent lineups

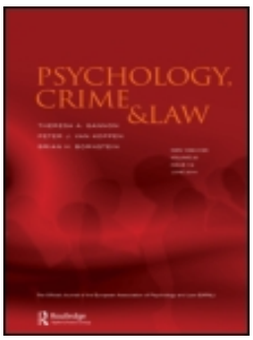
Calderwood, L.; McKay, D.R.; Stevenage, S.V.

Published in:
Psychology, Crime and Law

DOI:
10.1080/1068316X.2019.1597090

E-pub ahead of print: 23/04/2019

Document Version
Peer reviewed version



Routledge
Taylor & Francis Group

Psychology, Crime & Law

ISSN: 1068-316X (Print) 1477-2744 (Online) Journal homepage: https://www.tandfonline.com/loi/gpcl20

Children's identification of unfamiliar voices on both target-present and target-absent lineups

L. Calderwood, D. R. McKay & S.V. Stevenage

To cite this article: L. Calderwood, D. R. McKay & S.V. Stevenage (2019): Children's identification of unfamiliar voices on both target-present and target-absent lineups, Psychology, Crime & Law, DOI: 10.1080/1068316X.2019.1597090

To link to this article: https://doi.org/10.1080/1068316X.2019.1597090

Accepted author version posted online: 21 Mar 2019.

Publisher: Taylor & Francis & Informa UK Limited, trading as Taylor & Francis Group

Journal: *Psychology, Crime & Law*

DOI: 10.1080/1068316X.2019.1597090

Children's identification of unfamiliar voices on both target-present and target-absent lineups

Calderwood, L.[a]*, McKay, D. R.[a] and Stevenage, S.V.[b]

[a] *Department of Psychology, University of the West of Scotland, Paisley, Scotland;* [b]*Department of Psychology, University of Southampton, Southampton, England*

* Correspondence should be addressed to Dr Lesley Calderwood at:
School of Media, Culture and Society, University of the West of Scotland, Paisley Scotland, PA1 2BE
Email: Lesley.Calderwood@uws.ac.uk; Tel: 0141 848 3895

Children's identification of unfamiliar voices on both target-present and target-absent lineups

Abstract

A robust finding from the eyewitness literature is that children are as accurate as adults on target-present lineups from the age of five years, whereas they continue to make an erroneous false positive identification from a target-absent lineup up until fourteen years (Pozzulo & Lindsay, 1998). The current study explores whether the same pattern occurs when voices are used instead of faces and evaluates the reliability of children as potential earwitnesses. A total of 334 participants from six age groups (6-7-year-olds, 8-9-year-olds, 10-11-year-olds, 12-13-year-olds, 14-15-year-olds and adults) listened to a 30 second audio clip of an unfamiliar voice and were then presented with either a six person target-present or target-absent voice lineup. Overall, participants were more accurate with target-present than target-absent lineups. Performance on target-present lineups showed adult-like levels of attainment by 8-9 years of age. In contrast, performance on target-absent lineups was extremely poor and remained poor through to adulthood with all age groups tending to make a false identification. Confidence was higher when participants made correct than incorrect decisions for both types of lineup and this did not change with increasing age. Given these results, both child and adult earwitness evidence needs to be treated with considerable caution.

Introduction

In most criminal cases the perpetrator has been seen by a victim or eyewitnesses and their visual descriptions are used to aid the identification process. However, there are often cases in which the victim's or witness's memory of the perpetrator's voice can provide a useful clue to identification. There may even be occasions when the voice is the only clue to identification such as when crimes are committed in the dark, over the phone, or when the perpetrator is wearing a disguise or visibility is reduced. Whilst we know a lot about the reliability (and fallibility) of eyewitness identification, very little is known about the reliability of earwitness identification and even less is known about children's voice identification abilities. The purpose of the present paper is to examine children's voice identification abilities with a view to informing the question of whether they would be reliable earwitnesses in a criminal case.

Previous Research

Adult's voice recognition abilities

Research with adults has consistently shown that voice identification is much poorer than face identification (see Yarmey, 1995, for a review). For example, Öhman, Eriksson and Granhag (2012) tested adults' recognition of an unfamiliar voice using a seven voice lineup after a two week delay. Surprisingly, only 19% of adults could correctly identify the target voice. Although this result was significantly above chance, performance is poor when compared to unfamiliar face recognition. Adults also find it more difficult to recognise familiar people from their voice than their face. Hanley, Smith and Hadfield

(1998) asked participants to identify famous people from their face or their voice. They had to firstly say whether the person was familiar and if so, were then asked to provide semantic information to identify the person or state the person's name. Participants went on to name 73% of the faces they had found familiar but could only name 44% of the voices that were familiar. Using personally familiar people as opposed to celebrities, Yarmey, Yarmey, Yarmey, and Parliament, (2001), found that adults identified highly familiar voices with only 85% accuracy. It has been suggested that we can only get comparable performance on familiar face and voice recognition tasks when the faces are blurred (Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009). This has led to the proposal that there is a weaker route for voice identification whereby most of our attention is focused on processing the content of what is being said rather than on the identity of who is speaking or their emotional state (Goggin, Thompson, Strube, & Simental 1991). In summary, results from a range of different studies have consistently shown that adults find it much harder to recognise people from their voices than from their faces.

Children's voice recognition abilities

Against this backdrop, there is little research to date on children's recognition of voices. Consequently, we do not know whether they too show relatively poor performance with voices compared to faces. In a more applied sense, we are yet to understand how reliable they would be as earwitnesses. Early work shows that children's recognition of *familiar* voices, like familiar faces, is highly accurate. For example, when presented with a 4 second clip of a familiar cartoon voice and asked to point to the corresponding cartoon picture, young children performed surprisingly well (Spence, Rollins & Jerger, 2002). The youngest group of 3-year olds performed with 61% accuracy, whilst the 4 and 5 year

olds performed with 81% and 86% accuracy respectively. One of the problems with cartoon voices is that there is no guarantee that all children will know all of the voices. One way to resolve this has been to use personally familiar voices. Taking this approach, Murry and Cort, (1971), found high levels of accuracy for their classmates' voices in 9-10-year-olds. Thus, children's ability to recognise familiar voices follows a similar pattern to faces where familiar face recognition has reached adult-like levels by around 5-6 years of age (e.g., Diamond & Carey, 1977; Pozzulo, Dempsey, Bruer, & Sheahan, 2012).

In relation to the identification of *unfamiliar* voices, there are only a few studies that have been conducted with children, and the results are inconsistent regarding the stage at which adult-like levels of performance are obtained. Mann, Diamond and Carey (1979) examined the development of voice recognition skills between the ages of 6 and 16 years. Children listened to a short audio clip and were then presented with two voices and were asked to decide which of them belonged to the first voice that they had heard. Performance was better when the same utterance was heard at study and at test. However, of more interest was the developmental trajectory of voice recognition skills. In this regard, six-year-old children performed below chance levels, but performance increased between the ages of 6 and 10 with adult-like performance achieved by 10 years. A developmental dip was shown between 10 and 13 years of age, mirroring a similar dip in face recognition (Carey, Diamond & Woods, 1980). Following this dip, adult levels of voice recognition performance were again apparent by age 14.

To the authors' knowledge, the study by Mann et al. (1979) is the only one to have examined the development of voice matching skills across such a wide age range. Given

that this set of data are nearly 40 years old, the current study will test a similar age range of children (6-15-year-olds) and compare them with adults to establish whether a similar pattern of age-related improvement in voice recognition is found using a different task. Rather than present a target voice followed by two test voices as in Mann et al.'s (1979) matching study, the current study uses a lineup identification task in which children hear a target voice and then hear a selection of six voices. Their task is to identify whether one of these lineup voices matches the target voice. The advantage of such a task is the possibility to examine the potential for bias in children's responding, by exploring their performance on both target-present and target-absent trials.

Children's high rate of false identifications in target-absent lineups

It is of vital importance to include both target-present and target-absent lineups in this study as it has been consistently shown in the eyewitness literature that children reach adult levels of performance at different ages on each type of lineup. Adult-like levels of accuracy are consistently found on target-present lineups from the age of 5-6 years (see Pozzulo & Lindsay, 1998, for a review), however, adult-like performance on target-absent lineups is not achieved until 14-15 years of age. Instead of correctly responding that the target is not there on a target-absent lineup, children continue to pick someone from the lineup leading to a high number of false identifications in eyewitness tasks that does not improve with age. The real world importance and possible consequence of erroneous false positive identifications by children should not be underestimated and therefore increasing knowledge of this issue is of the utmost importance.

In a recent review of this area by Havard (2014), it is suggested that children's high levels of false identifications in target-absent lineups are largely driven by a social pressure to pick someone from the lineup to please the experimenter rather than due to immature face processing abilities. This is similar to the criticisms made of the Piagetian conservation tasks, where children are repeatedly asked the same question and so think that they should give a different answer (Rose & Blank, 1974). In support of this view, it has been found that when the target-absent lineup contains a 'mystery man' or a silhouette that children can pick, false identifications are greatly reduced (Havard & Memon, 2012; Zajac & Karageorge, 2009). Dunlevy and Cherryman (2013) included a tree in target-present and target-absent lineups and told child participants that they should select the tree if they thought the person was not there and was hiding behind the tree. The provision of this 'tree' option reduced false positive identifications dramatically in 6-7-year-old children and is consistent with the theory that children's high false positive rates may reflect a lack of complete understanding of the task, and an implicit desire to choose.

The current study will provide further insight into this problematic pattern of responding by using a different stimulus, voices, to investigate whether the high level of false identifications extends to a different stimulus than faces. If this is the case, then the results will lend support to the proposal that children are largely driven by an implicit pressure to pick someone in a target-absent lineup and that their poor performance on these lineups is not due to immature face processing abilities. These results will therefore have important implications for those working with both child eyewitnesses and earwitnesses.

Children's performance on voice lineups

In recent years, a few earwitness studies have been carried out using a lineup paradigm with children, but only one of these used both target-present and target-absent lineups. Öhman, Eriksson and Granhag, (2011) tested 7-9-year-olds, 11-13-year-olds and adults. Participants listened to a 40 second voice clip before making an identification decision from a 7-person lineup two weeks later. Half of the participants were given a target-present lineup and the remainder were given a target-absent lineup. For target-present trials, the 11-13-year-olds performed better than both the younger children and the adults, and were the only group to identify the target at a level above chance. For the target-absent trials, both adults and children did not differ in making a high level of false identifications (60% vs 49% respectively). Despite this, both children and adults performed better on target-absent than target-present lineups, which is the opposite pattern to the eyewitness literature. With only a single study examining performance in both target-present and target-absent voice lineups in which only one age group performed above chance on the target-present lineups, more research is essential if a reliable picture of performance is to be obtained and will help to establish whether the opposite pattern of responding is found for faces and voices.

The Present Study

Based on the previous review, the present study will employ both target-present and target-absent trials in a voice lineup paradigm. In extension to the previous work, multiple target voices will be used so that results may be generalised beyond the single

target voice used in previous studies. To avoid floor effects, testing will be conducted immediately rather than after a delay, and the methodology will allow the presentation of a long speech clip at study as well as the opportunity to listen to the test voices twice before making a decision. On the basis of previous work, it was expected that performance would improve on target-present and target-absent lineups with increasing age, reaching adult levels by about 10 years. It was also expected that performance may be better on target-absent than target-present lineups across all age groups.

Method

Design

A 6 x 2 between-groups design was used in which voice identification was investigated across six age groups (6-7-year-olds, 8-9-year-olds, 10-11-year-olds, 12-13-year-olds, 14-15-year- olds, and adults) on both target-present and target-absent lineups. Accuracy and confidence were the dependent variables.

Participants

A total of 334 participants took part in the present study, with 163 participants receiving a target-present lineup and 171 receiving a target-absent lineup. All were drawn from the West of Scotland area and thus were familiar with the accent of the speakers whilst being unfamiliar with the speakers themselves. This project was approved by the School of Media, Culture & Society Ethics Committee at the University of the West of Scotland and in accordance with this approval, written parental consent was obtained for all child participants. Verbal assent was also obtained on the day of testing from all child participants. Adult participants provided their own written consent to take part in the study. Participants were randomly allocated to the lineup condition with similar numbers in each age group as detailed in Table 1.

Insert Table 1 around here

Materials

Two speech clips were obtained from a total of 24 Scottish female speakers, aged between 25-35 years, recruited from the West of Scotland area. All had a standard accent typical of the region and were free from speech impediments. The voice clips contained free rather than scripted speech. To obtain this, participants were shown a picture of two scenes, a farmyard and a fairground, and were asked to describe them. From these recordings, a thirty second clip of one scene was selected for the study phase and an eight second clip of the other scene was selected for the lineup phase. In this way, the content differed between the study and lineup phases and so could not be used to help with the identification of the target speaker.

From this database of 24 speakers, the ratings from nine local participants were used to select target speakers, target-replacements (for target-absent lineups), and the foils. Two target speakers were selected on the basis that they had no distinguishing characteristics such as pitch, speaking rate and modularity. The remaining 22 speakers were compared to the targets and the most similar voice for each target (as rated on a 5-point scale) was selected as the target-replacement for the target-absent lineups. The next five most similar voices for each target were selected as the foils for both the target-present and target-absent lineups. Analysis of the ratings confirmed that there was no significant difference in the perceived similarity of the two target replacement voices used in the target-absent lineups to their respective target voice (target 1 similarity = 3.11, target 2 similarity = 3.22, $t(8) = 0.32$, $p>0.05$). The remaining voices were not used in the experiment.

Audacity 1.3 Beta was used to edit the clips of the targets, the foils and the target-replacements to produce a thirty second clip of the two targets describing the photograph

of the farm for the study phase, and an eight second clip of each speaker describing the fairground for the lineup phase. These timings were in line with previous work using a similar paradigm (e.g., Stevenage, Clarke & McNeill, 2012; Öhman et al., 2012).

From these stimuli, four lineups were created: a target-present and a target-absent lineup for Speaker 1 and a target-present and a target-absent lineup for Speaker 2. Each participant heard one lineup and care was taken to vary the position of the target, or the target-replacement in each lineup for each participant.

All the voices were presented through a PowerPoint presentation which participants played at their own pace. Testing took place in a quiet environment, however, headphones were used to minimise distraction from ambient noise.

Procedure

Both child and adult participants were tested in small groups of four or five within a quiet area of their school or University. Each participant completed the lineup using a laptop and a set of headphones. They were instructed to listen carefully to the target voice and they were encouraged to focus on the voice rather than the content in readiness for a recognition test later. Each target voice was then played for thirty seconds. Following this, participants were told that they would hear six short clips and they would be asked to consider whether any of them sounded like the target. They were warned that the target may not be one of the six voices and that half of the lineups were target-present and half of them were target-absent. After listening to each of the six clips, participants were then told that they would hear the six clips again and that this time they had to make a decision

for each voice about whether it was the same person as the target voice. For each voice, they marked either Yes or No on a sheet of paper beside them. The lineup task was completed either when all six test voices had been heard, or when a Yes decision had been indicated. Participants were then asked to rate how confident they felt about their decision on a scale of 1-5, with 1 indicating a guess and 5 indicating that they were very sure.

Results

Accuracy

In terms of accuracy, participants listening to a target-present lineup could respond with a hit (correct identification of the target), false identification (selection of a foil) or a miss (incorrectly saying that the target was not there). Similarly, participants listening to a target-absent lineup could respond with a correct rejection (correctly saying that the target was not there) or a false identification (selecting the target replacement or any one of the foils). The pattern of performance for each age group is presented and analysed separately for target-present and target-absent lineups in order to determine whether there are different patterns of responding for each type of lineup.

Target-present lineups

Figure 1 shows the percentage of responses for each possible response for each age group. As shown in Figure 1, 6-7-year-olds were more likely to choose someone else in the lineup and make a false identification, whereas all other age groups were more likely to

make a correct identification. Miss responses (rejecting the lineup and saying that the target was not there) were uncommon in the two older groups and the adult group, while the other groups, in particular, the 10-11-year-olds, did make a fair number of incorrect rejections.

Insert Figure 1 around here

A chi-square test was used to explore the association between age (6-7, 8-9, 10-11, 12-13, 14-15-year-olds and adults) and response type (correct identification, false identification and miss). Given that there were several cells with an expected count of less than 5, Fisher's Exact test was used. This confirmed a significant association between age and response ($p<0.001$). Analysis of the standardised residuals indicated that this was due to a particularly low level of performance in the 6-7 year olds, together with a particularly good level of performance as indicated by very few 'miss' decisions from 12-13 year olds. This is suggestive of Mann et al.'s (1979) demonstration of the attainment of adult levels of performance by the age of 10 followed by a developmental dip and then recovery of performance levels thereafter. However, the adult level of attainment was observed by 8-9 years of age rather than by 10 years. Indeed, when the 6-year-olds were removed from the analysis, Fisher's Exact test showed there was no significant association between age and response, $p=0.22$, suggesting that adult level of performance is attained by 8-9 years of age and no further improvement in performance is observed.

Target-absent lineups

As can be seen in Figure 2, the rate of false identifications in target-absent lineups was high and remained high across all age groups.

Insert Figure 2 around here

As above, a chi-square test was used to explore the association between age (6-7, 8-9, 10-11, 12-13, 14-15-year-olds and adults) and response type (correct rejection, false identification). This revealed no significant association between the two variables, $\chi^2(5, N = 171) = 3.37, p=0.65$. In fact, all age groups were more likely to make a false identification and pick someone from the lineup than correctly saying that the person was not there.

Overall, the results suggested some improvement in performance with target-present lineups, when 6-7 year olds were compared to older participants. The absence of any association between age and response after this age group suggested that adult levels of performance had been attained by 8-9 years of age. In contrast, no age-related improvements emerged with target-absent lineups. In fact, performance was rather poor in target-absent lineups compared to target-present lineups through the inappropriate tendency to select a voice rather than indicate that the target was not present.

Confidence

Once participants had made their decision, they were asked to rate how confident they were on a scale from 1-5 (where 1 indicated that they were guessing and 5 indicated that they were very confident that they were correct).

Target-present lineups

The mean confidence scores for correct and incorrect responses on target-present lineups are presented below in Figure 3. Confidence tended to be higher for most age groups when responses were correct, however, the 6-7-year-olds and the 14-15-year-olds were slightly more confident when incorrect.

Insert Figure 3 around here

A 6 (age group) x 2 (lineup accuracy: correct vs incorrect) between-groups Analysis of Variance (ANOVA) was conducted on the confidence data. This revealed a main effect of lineup accuracy only, with confidence being higher when correct (M= 4.11) than when incorrect (M=3.63), $F_{(1, 149)} = 5.01$, $p = 0.027$, $\eta^2 p = 0.03$. There was no significant main effect of age, $F_{(5, 149)} = 0.699$, $p = 0.625$ and no significant interaction between age and accuracy, $F_{(5, 149)} = 0.96$, $p = 0.44$ suggesting that the improvement in accuracy with age was not mirrored by a similar increase in confidence.

Target-absent lineups

The mean confidence scores for correct and incorrect responses on target-absent lineups are presented below in Figure 4. Confidence was higher for all age groups when responses were correct.

Insert Figure 4 around here

For target-absent lineups, a 6 x 2 between groups ANOVA was again conducted. The results mirrored those with target-present lineups in all respects. Specifically, a significant main effect of lineup accuracy emerged, with higher confidence when correct (M=3.98) than when incorrect (M=3.57), $F_{(1, 151)} = 4.71$, $p = 0.032$, $\eta^2 p = 0.03$. As before, there was no significant main effect of age, $F_{(5, 151)} = 0.41$, $p = 0.84$, and no significant interaction, $F_{(5, 151)} = 0.19$, $p = 0.97$.

In summary, for both target-present and target-absent lineups confidence was significantly higher when participants' decisions were correct than when incorrect regardless of age group. In the target-absent case, stable confidence levels mirrored stable accuracy levels. Interestingly, however, in the target-present case, improvements in accuracy with age were not accompanied by an increase in confidence. This said, in both cases, the effect sizes were small and the interpretation of confidence data should therefore be treated cautiously.

Discussion

The aim of this study was to examine children's voice matching performance with a view to investigating the developmental pattern shown by Mann et al. (1979) using a different paradigm. In this regard, the results did not entirely support the previous pattern of development across the age range. Rather than showing attainment of adult levels of performance by 10 years of age, the analysis of target-present performance actually indicated adult levels of attainment in slightly younger participants of 8-9 years of age. There was also some evidence of a developmental dip in the 10-11 year olds, with a significant recovery in performance by 12-13 year olds, as shown by the absence of any 'misses' when the target was present in the lineup. This result sits well alongside that of Öhman et al. (2012) who showed no improvement in target-present voice performance when 11-13 year olds were compared with adults.

When examining performance on target-absent lineups, the lack of any improvement with age was clear. In fact, all participants showed equivalent and rather poor performance with target-absent lineups compared to target-present lineups. Performance in the target-absent case was marked by a substantial tendency to select a voice inappropriately from the lineup.

Together, these results sit at odds with those obtained from face lineups where both target-present performance (e.g., Bruce et al, 2000; Megreya & Bindemann, 2015) and target-absent performance (e.g., Pozzulo & Lindsay, 1998; Havard, 2014) have been shown to improve with age. This discrepancy is unlikely to be due to floor effects, as the use of a long (30 second) clip at study, and the use of an immediate test, ensured that performance exceeded that of previous studies (c.f., Öhman et al., 2012). Instead, the current results

underline the difficulty of the voice matching task, immediate or otherwise, compared to a face matching task.

Target-Absent Performance

The particularly poor performance in the target-absent condition, relative to the target-present condition, was surprising given the opposite pattern in Öhman et al.'s (2011) study. It is also of particular applied interest given the real-world consequences of inappropriately selecting from a target-absent lineup. It has been suggested that children's high false positive identifications may be driven largely by a social pressure to choose and that false identifications dramatically reduce when a mystery man, a silhouette or a tree is included in the lineup for children to pick when they think that the person is not there (see Havard, 2014, for a review). The provision of these options across different studies reduced false positive identifications dramatically in children and is consistent with the theory that children's high false positive rates may reflect a lack of complete understanding of the task, and an implicit desire to choose.

The same may also be true of adults. Indeed, in a study carried out by Van Wallendael, Surace, Parsons, and Brown, (1994), participants listened to a sales pitch and then were given a target-present or target-absent lineup either immediately, or after 7 or 14 days. On the target-present lineups, correct identification rates were 80% when tested immediately, 90% after the 7 day interval and 76% after the 14 day interval. In contrast, on target-absent trials, all bar one participant (out of 76) made a false identification, regardless of the length of the delay. Similarly, using the same procedure as the current

study, Stevenage et al., (2012), found that adults performed more poorly on target-absent lineups (34% correct rejection rate) than target-present lineups (57% correct identifications), (see also, Philippon, Cherryman, Bull & Vrij, 2007). We have therefore found the same pattern of results as reported in eyewitness paradigms where children perform dramatically better in target-present trials compared to target-absent trials. Further, in the case of adults, we have found that they too perform much better in target-present trials and make a significantly large number of false identifications when listening to voices. It seems to be the case that, unlike faces, false identifications for voices are high in childhood and remain high throughout adulthood. Matching identities for voices is a difficult task for children and remains difficult through to adulthood. This finding has serious implications for those working with earwitnesses as there was very little delay between hearing the initial voice and making the lineup identification.

Several explanations may exist to account for the high rate of false identifications in target-absent trials for voices. For instance, it is possible that the lineups had been created such that the similarity between target and foil voices was too high to support effective target identification. Alternatively, the memory load associated with a necessarily sequential voice lineup created task demands that were too difficult. In both cases, however, it is difficult to see why the performance in target-absent trials was so much worse than that in target-present trials where the same issues existed.

Perhaps more fruitful is a consideration of interference effects when recognising voices (Stevenage et al., 2013) in which the presentation of intervening voices between study and test can dramatically impair the memory for the original voice. This may have the capacity to account for poorer performance in target-absent trials than in target present

trials, as the lack of the target means that every voice in the target absent lineup is a distractor voice.

Alongside this, the eyewitness literature points to an implicit social pressure to choose one of the alternatives when presented with a lineup. This well-known concern is usually addressed through a reminder to the participants that the 'target may or may not be present in the lineup'. However, despite this, the pressure to choose may still exist. The capacity to choose the 'tree' in Dunlevy and Cherryman's (2013) study provides a lovely solution to address this concern for children, and it would be interesting to explore whether the provision of a positive option as a way for adults to indicate that the target is not present may similarly reduce false identification rates in voice lineups such as a scrambled voice that participants could choose if they felt that the target voice was not in the lineup.

A Reflection on Confidence

In turning to a consideration of confidence ratings, the value of witness confidence ratings has been debated for many years. In the eyewitness literature, there traditionally appeared to be a weak positive relationship at best between confidence and accuracy and confidence was generally not regarded as a reliable predictor of identification accuracy. A similar observation has been made regarding confidence in the earwitness literature, with early studies showing a low or non-significant relationship often reported for voice identification in adults (Yarmey, 1995; Yarmey, 2001). Öhman et al., (2011), found no evidence for a relationship between confidence and accuracy for their child and adult participants and, whilst Öhman et al., (2013), did not formally measure confidence, when they asked participants if they thought they would be able to recognise a target voice,

86% of children and 63% of adults responded in the positive. In reality, only 13% of children and 4% of adults could correctly identify the target voice. Within the current study, confidence emerged as being significantly higher when participant decisions were correct than when incorrect, across both target-present and target-absent lineups. This pattern held regardless of participant age suggesting that participants knew when they were right suggesting that confidence could be a useful indicator of accuracy in earwitness studies.

This is in line with more recent thinking on the relationship between confidence and accuracy in the eyewitness literature which has suggested that confidence may be a more reliable indicator of accuracy than originally thought (Wells, Olson & Charman, 2002). A meta-analysis by Sporer, Penrod, Read & Cutler, (1995), found that when those who chose someone from the lineup and those who did not choose someone from the lineup were analysed separately, the confidence of choosers gave a more reliable indication of accuracy than when all participants were analysed together. It is also argued that a confidence rating obtained at the time of the identification is a much more accurate indicator of accuracy than ratings obtained sometime after the initial identification (Brewer & Palmer, 2010). Calibration studies have also shown that confidence can be a more reliable indicator of accuracy than traditionally believed when measured this way than the traditional methods using correlation (Juslin, Olsson & Winman, 1996). Therefore, when eyewitnesses are tested using appropriate identification procedures, it is argued that the confidence they express can be a more reliable indicator of accuracy than was initially believed (Wixted & Wells, 2017). The current results suggest that confidence may also be a useful indicator of earwitness accuracy, however, these effects were

associated with very small effect sizes, and caution should be encouraged when assessing the reliability of confident earwitnesses.

Limitations and Future Work

One point worth reflecting on within the current study was the fact that all voices were obtained from adult speakers. Given own-age effects when recognising faces (Bonner & Burton, 2004; Rhodes & Anastasi, 2012), this may represent a confound in the current design. More specifically, it may have been more appropriate to test each age-group with voices drawn from their own age. This said, the design of the present study, whilst perhaps not optimal, would have biased performance in favour of the adult participants. In this regard, the fact that all bar the youngest children performed at a level comparable to the adults here is perhaps notable. Nevertheless, future work would be well-directed to test voice recognition across the age range by using age-relevant voices.

Conclusion

In conclusion, the present results show that voice recognition, even on immediate testing, was a rather difficult task. Performance was better in target-present than target-absent lineups, possibly reflecting a tendency to make a positive selection from the lineup rather than report that the target was not there. In target-present lineups, adult levels of attainment appeared to be demonstrated by 8-9 years of age, and recovered by 12-13 years of age after a slight developmental dip. Of particular concern was the very poor performance on target-absent lineup trials, with real-world implications being felt for the

innocent police suspect who may be selected from the lineup inappropriately. Given these results, both child and adult earwitness evidence needs to be treated with considerable caution. Even if a witness may have confidence in their identification, the current data suggests that earwitness performance may be too poor to rely on in court.

References

Bonner, L., & Burton, A. M. (2004). 7–11-year-old children show an advantage for matching and recognizing the internal features of familiar faces: Evidence against a developmental shift. *The Quarterly Journal of Experimental Psychology Section A*, *57*(6), 1019-1029.

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology*, *15*(1), 77-96.

Bruce, V., Campbell, R. N., Doherty-Sneddon, G., Langton, S., McAuley, S., & Wright, R. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology*, *18*(3), 319-333.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology*, *16*(4), 257.

Damjanovic, L., & Hanley, J. R. (2007). Recalling episodic and semantic information about famous faces and voices. *Memory & Cognition*, *35*(6), 1205-1210.

Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. *Journal of Experimental Child Psychology*, *23*(1), 1-22.

Dunlevy, J. R., & Cherryman, J. (2013). Target-absent eyewitness identification line-ups: Why do children like to choose. *Psychiatry, Psychology and Law*, *20*(2), 284-293.

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The Role of Language Familiarity in Voice Identification. *Memory and Cognition, 19,* 448–458.

Hanley, J. R., & Damjanovic, L. (2009). It is more difficult to retrieve a familiar person's name and occupation from their voice than from their blurred face. *Memory*, *17*(8), 830-839.

Hanley, J. R., Smith, S. T., & Hadfield, J. (1998). I recognise you but I can't place you: An investigation of familiar-only experiences during tests of voice and face recognition. *The Quarterly Journal of Experimental Psychology: Section A*, *51*(1), 179-195.

Havard, C. (2014). Are children less reliable at making visual identifications than adults? A review. *Psychology, Crime & Law*, *20*(4), 372-388.

Havard, C., & Memon, A. (2013). The mystery man can help reduce false identification for child witnesses: Evidence from video line-ups. *Applied Cognitive Psychology*, *27*(1), 50-59.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1304.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology*, *27*(1), 153-165.

Megreya, A. M., & Bindemann, M. (2015). Developmental improvement and age-related decline in unfamiliar face matching. *Perception*, *44*(1), 5-22.

Murry, T., & Cort, S. (1971). Aural identification of children's voices. *Journal of Auditory Research*, *11(3),* 260-262.

Öhman, L., Eriksson, A., & Granhag, P. A. (2011). Overhearing the planning of a crime: Do adults outperform children as earwitnesses? *Journal of Police and Criminal Psychology*, *26*(2), 118-127.

Öhman, L., Eriksson, A., & Granhag, P. A. (2012). Enhancing Adults' and Children's Earwitness Memory: Examining Three Types of Interviews. *Psychiatry, Psychology and Law*, *20*(2), 216-229.

Öhman, L., Eriksson, A., & Granhag, P. A. (2013). Angry Voices from the Past and Present: Effects on Adults' and Children's Earwitness Memory. *Journal of Investigative Psychology and Offender Profiling*, *10*(1), 57-70.

Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures. *Applied Cognitive Psychology*, *21*(4), 539-550.

Pozzulo, J. D., Dempsey, J., Bruer, K., & Sheahan, C. (2012). The culprit in target-absent lineups: Understanding young children's false positive responding. *Journal of Police and Criminal Psychology*, *27*(1), 55-62.

Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: a meta-analysis. *Law and Human Behavior, 22(5),* 549-570.

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: a meta-analytic and theoretical review. *Psychological Bulletin, 138(1),* 146-174.

Rose, S.A., & Blank, M. (1974). The potency of context in children's cognition: An illustration through conservation. *Child Development. 45(2),* 499–502.

Spence, M. J., Rollins, P. R., & Jerger, S. (2002). Children's recognition of cartoon voices. *Journal of Speech, Language, and Hearing Research*, *45*(1), 214-222.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315.

Stevenage, S. V., Clarke, G. & McNeill, A. (2012). The "other-accent" effect in voice recognition. *Journal of Cognitive Psychology, 24,* 647-653.

Stevenage, S. V., Neil, G. J., Barlow, J., Dyson, A., Eaton-Brown, C., & Parsons, B. (2013). The effect of distraction on face and voice recognition. *Psychological research*, *77*(2), 167-175.

Van Wallendael, L. R., Surace, A., Parsons, D. H., & Brown, M. (1994). 'Earwitness' voice recognition: Factors affecting accuracy and impact on jurors. *Applied Cognitive Psychology*, *8*(7), 661-677.

Wells, G. L., Olson, E. A., & Charman, S. D. (2002). The Confidence of Eyewitnesses in Their Identifications From Lineups. *Current Directions in Psychological Science, 11,* 151-154.

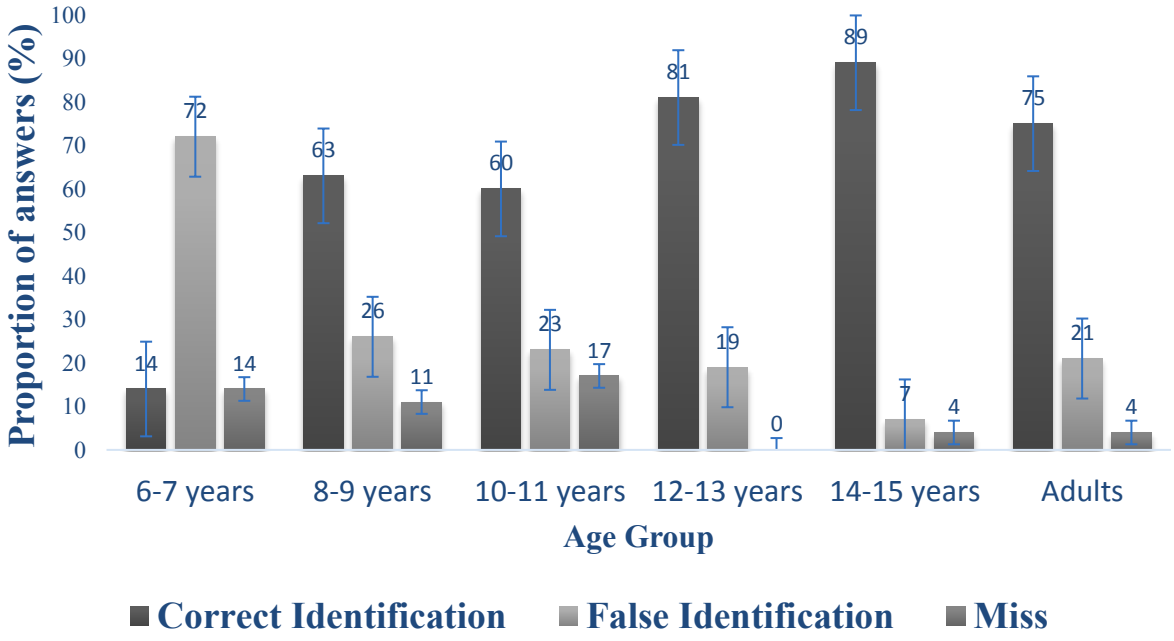
Wixted. J. T. & Wells, G. L. (2017). The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis. *Psychological Science in the Public Interest,* 18, 10–65.

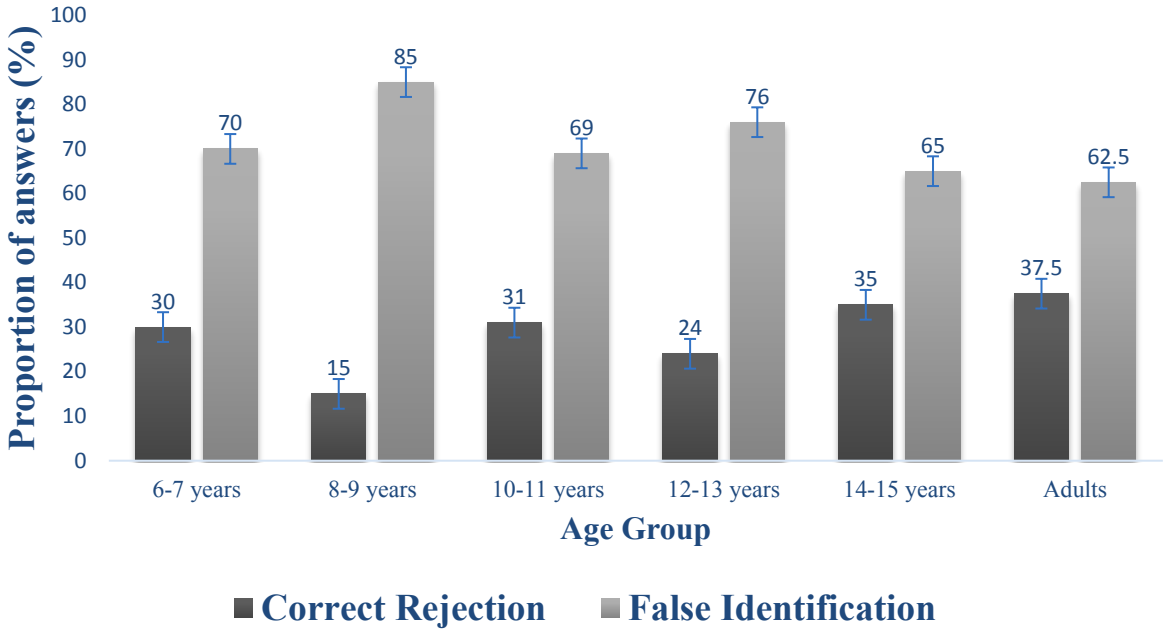
Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law*, *1*(4), 792.

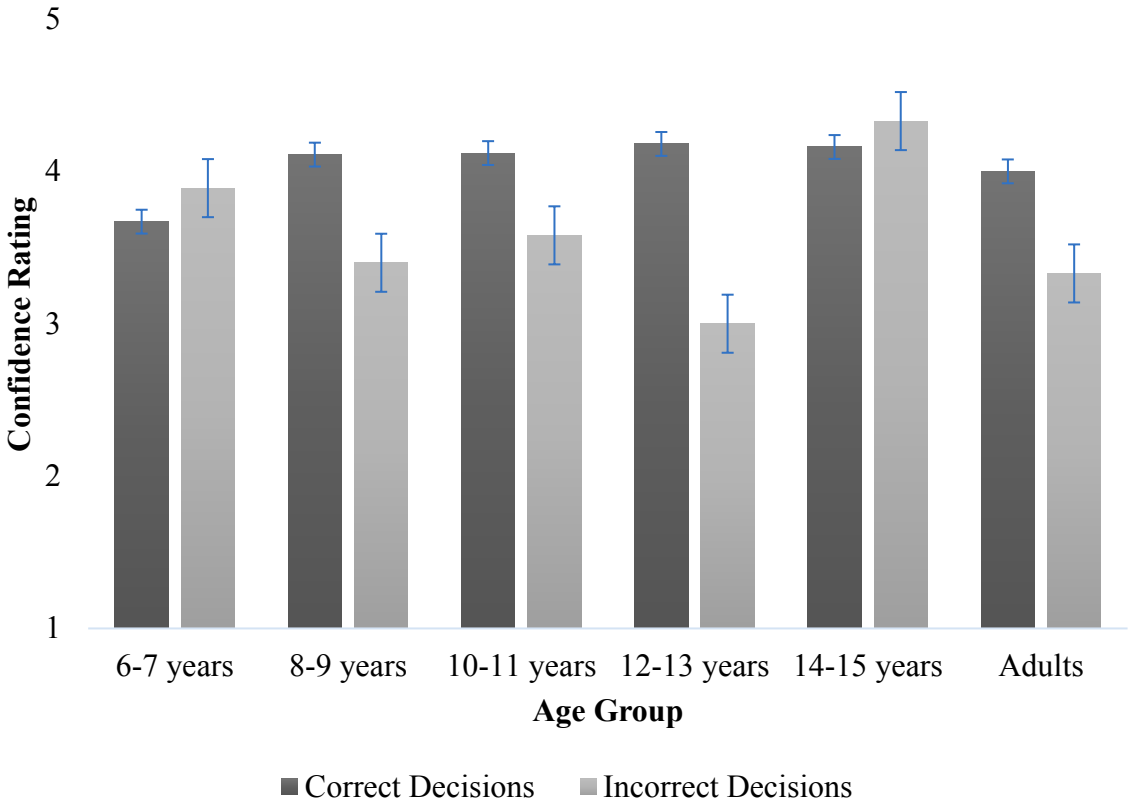
Yarmey, A. D. (2001). Expert testimony: Does eyewitness memory research have probative value for the courts? *Canadian Psychology*, *42*(2), 92-

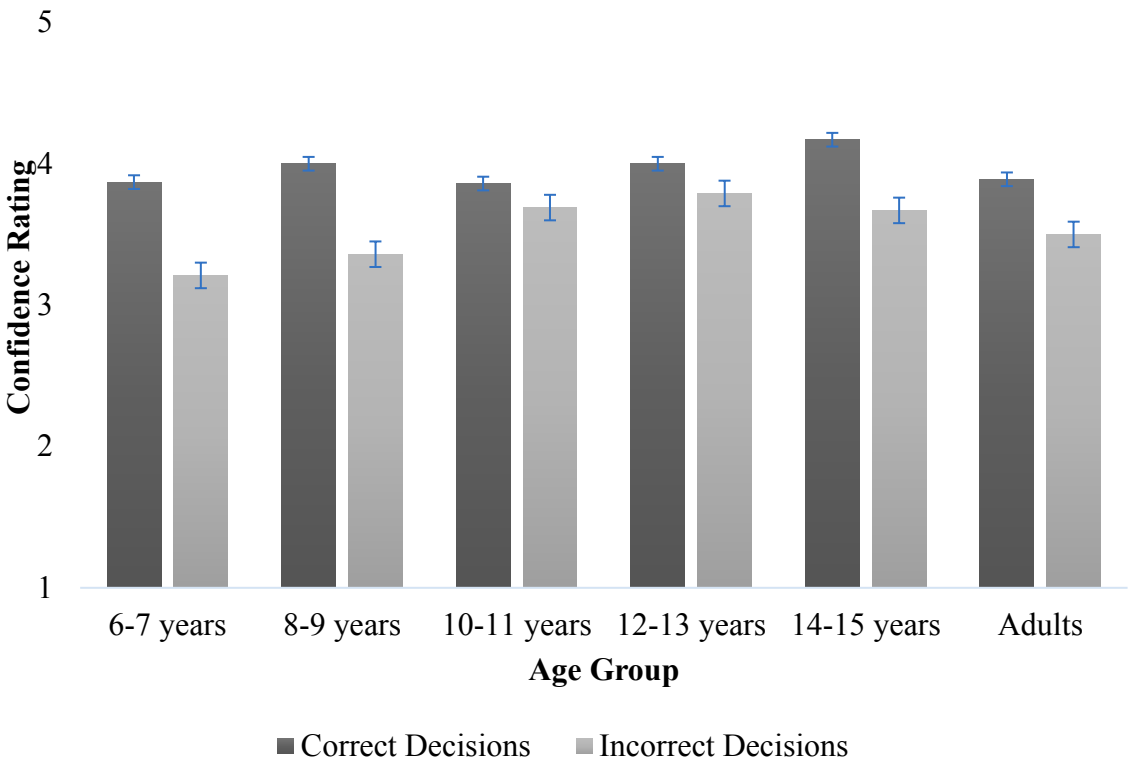
Yarmey, A. D., Yarmey, A. L., Yarmey, M. J., & Parliament, L. (2001). Commonsense beliefs and the identification of familiar voices. *Applied Cognitive Psychology*, *15*(3), 283-299.

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology*, *23*(3), 358–368.


Proportion of answers (%)
100
90
80
70
60
50
40
30
20
10
0
14
72
14
63
26
11
60
23
17
81
19
0
89
7
4
75
21
4
6-7 years
8-9 years
10-11 years
12-13 years
14-15 years
Adults
Age Group
Correct Identification
False Identification
Miss


Proportion of answers (%)
100
90
80
70
60
50
40
30
20
10
0
30
70
15
85
31
69
24
76
35
65
37.5
62.5
6-7 years
8-9 years
10-11 years
12-13 years
14-15 years
Adults
Age Group
Correct Rejection
False Identification


5
4
3
2
1
Confidence Rating
6-7 years
8-9 years
10-11 years
12-13 years
14-15 years
Adults
Age Group
Correct Decisions
Incorrect Decisions


5
4
3
2
1
Confidence Rating
6-7 years
8-9 years
10-11 years
12-13 years
14-15 years
Adults
Age Group
Correct Decisions
Incorrect Decisions

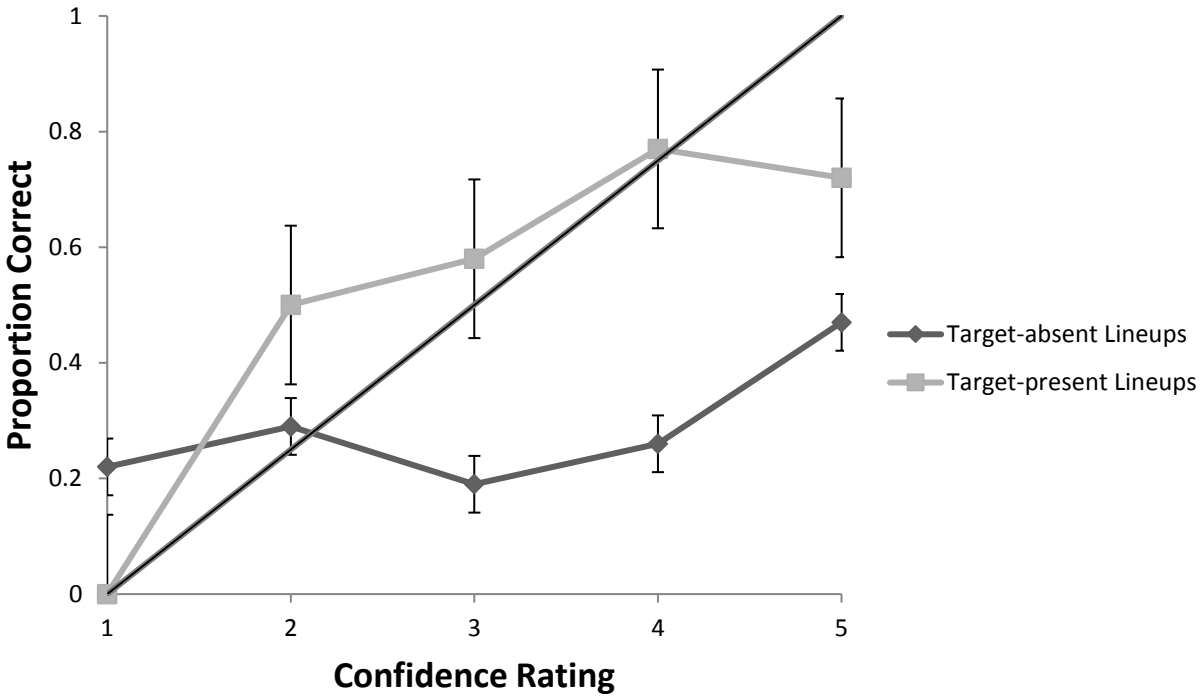

Proportion Correct
1
0.8
0.6
0.4
0.2
0
1
2
3
4
5
Confidence Rating
Target-absent Lineups
Target-present Lineups

Table 1 shows participant characteristics for each age group

Age group	**Mean Age**	**Gender**
6-7-year-olds (N= 44)	6 years, 8 months	24 F, 20 M
	Range (6, 0 – 7,8)	
8-9-year-olds (N= 54)	9 years, 0 months	26 F, 28 M
	Range (8,0 – 9, 9)	
10-11-year-olds (N= 56)	10 years, 7 months	30 F, 26 M
	Range (10, 0 – 11, 9)	
12-13-year-olds (N= 70)	12 years, 5 months	36 F, 34 M
	Range (12, 0 – 13,3)	
14-15-year-olds (N= 62)	14 years, 7 months	33 F, 29 M
	Range 14,0 – 15, 3	
Adults (N= 48)	28.84 years, SD = 7.97	25 F, 23 M
	Range (18-54 years)	